

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2000-311168
(P2000-311168A)

(43) 公開日 平成12年11月7日 (2000.11.7)

(51) Int.Cl.⁷
G 0 6 F 17/27
17/22

識別記号

F I
G 0 6 F 15/38
15/20

テ-リ-ド* (参考)

E 5 B 0 0 9
5 1 4 A 5 B 0 9 1

審査請求 有 請求項の数9 OL (全 9 頁)

(21) 出願番号 特願平11-123178

(22) 出願日 平成11年4月28日 (1999.4.28)

(71) 出願人 000004237

日本電気株式会社

東京都港区芝五丁目7番1号

(72) 発明者 玉野 隆一

東京都港区芝五丁目7番1号 日本電気株式会社内

(74) 代理人 100108578

弁理士 高橋 昭男 (外3名)

Fターム(参考) 5B009 MB03 MB22 TA06

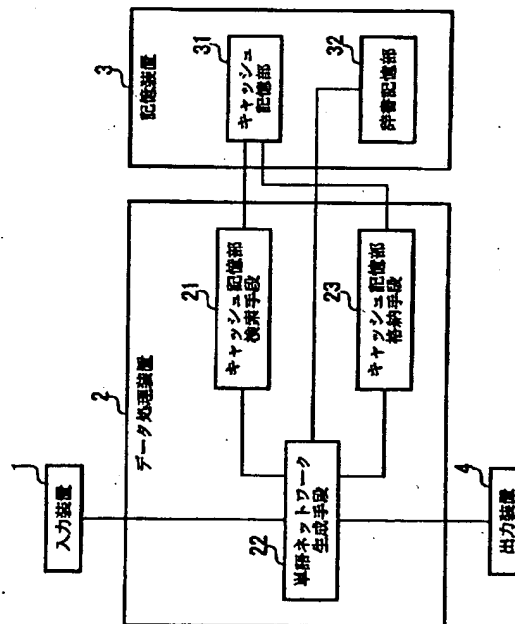
5B091 AA15 CA02 CC02 CD03 EA10

(54) 【発明の名称】 形態素解析システム及びその方法並びにこの形態素解析プログラムを記録した記録媒体

(57) 【要約】

【課題】 単語辞書のすべての単語ネットワークと、分割される可能性のある単語の全てとを比較することなく、形態素解析システム用の候補単語ネットワークを高速に生成する形態素解析システム及びその方法を提供する。

【解決手段】 形態素解析システムは、キーボード等の入力装置1と、プログラム制御により動作するデータ処理装置2と、情報を記憶する記憶装置3と、ディスプレイ装置や印刷装置等の出力装置4とを含んで構成されている。記憶装置3は、キャッシュ記憶部31と、辞書記憶部32とを備えている。キャッシュ記憶部31は、単語ネットワーク生成手段22で処理された単語ネットワークの中から、キャッシュ記憶部格納手段23により検出された単語ネットワークを逐次格納する。辞書記憶部32は、日本語の単語辞書であり、日本語の単語と品詞情報とをあらかじめ記憶している。



【特許請求の範囲】

【請求項1】 入力された文字列を単語に分割する形態素解析システムにおいて、
単語及び品詞情報を記憶している辞書記憶部と、
この単語記憶部を検索し、前記文字列を単語に分割し、
単語ネットワークを生成する単語ネットワーク生成手段と、
この単語ネットワーク生成手段により生成された単語ネットワークから、部分単語ネットワークを検出し、キャッシュ記憶部へ格納するキャッシュ記憶格納手段とを具備することを特徴とする形態素解析システム。

【請求項2】 入力される文字列を単語に分解するとき、既に解析された単語ネットワークが前記キャッシュ記憶部にある場合、このキャッシュ記憶部から取得し、解析結果の単語とするキャッシュ記憶部検索手段を具備することを特徴とする請求項1記載の形態素解析システム。

【請求項3】 キャッシュ記憶格納手段が入力された文字列を単語に分解した後、前記単語ネットワークの1単語に収束する部分を検出し、この収束する部分で入力文字列を区切り、部分単語ネットワークを構成し、各部分単語ネットワークを前記キャッシュ記憶部に格納することを特徴とする請求項1または請求項2記載の形態素解析システム。

【請求項4】 入力された文字列を単語に分割する形態素解析方法において、
単語及び品詞情報を辞書記憶部へ記憶させる辞書記憶過程と、
この単語記憶部を検索し、前記文字列を単語に分割し、
単語ネットワークを生成する単語ネットワーク生成過程と、
この単語ネットワーク生成過程により生成された単語ネットワークから、部分単語ネットワークを検出し、キャッシュ記憶部へ格納するキャッシュ記憶格納過程とを有することを特徴とする形態素解析方法。

【請求項5】 入力される文字列を単語に分解するとき、既に解析された単語ネットワークが前記キャッシュ記憶部にある場合、このキャッシュ記憶部から取得し、解析結果の単語とするキャッシュ記憶部検索過程を有することを特徴とする請求項4記載の形態素解析方法。

【請求項6】 キャッシュ記憶格納過程が入力された文字列を単語に分解した後、前記単語ネットワークの1単語に収束する部分を検出し、この収束する部分で入力文字列を区切り、部分単語ネットワークを構成し、各部分単語ネットワークを前記キャッシュ記憶部に格納することを特徴とする請求項4または請求項5記載の形態素解析システム。

【請求項7】 入力された文字列を単語に分割する処理をコンピュータに行わせる形態素解析プログラムを記録した記録媒体において、

単語及び品詞情報を辞書記憶部へ記憶させる辞書記憶処理と、

この単語記憶部を検索し、前記文字列を単語に分割し、
単語ネットワークを生成する単語ネットワーク生成処理と、

この単語ネットワーク生成処理により生成された単語ネットワークから、部分単語ネットワークを検出し、キャッシュ記憶部へ格納するキャッシュ記憶格納処理とをコンピュータに行わせることを特徴とする形態素解析プログラムを記録した記録媒体。

【請求項8】 入力される文字列を単語に分解するとき、既に解析された単語ネットワークが前記キャッシュ記憶部にある場合、このキャッシュ記憶部から取得し、解析結果の単語とするキャッシュ記憶部検索処理を有することを特徴とする請求項7記載の形態素解析プログラムを記録した記録媒体。

【請求項9】 キャッシュ記憶格納処理が入力された文字列を単語に分解した後、前記単語ネットワークの1単語に収束する部分を検出し、この収束する部分で入力文字列を区切り、部分単語ネットワークを構成し、各部分単語ネットワークを前記キャッシュ記憶部に格納することを特徴とする請求項7または請求項8記載の形態素解析プログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、文字列として入力された日本語文の形態情報を出力する日本語の形態素解析システム及びその方法に係わるものである。

【0002】

【従来の技術】例えば、仮名文字変換や機械翻訳など、日本語の処理を実行するにあたっては、形態素解析と呼ばれる処理を行う必要がある。従来、形態素解析は、単語をキーとしてその語彙情報を記憶した辞書を検索しながら、文字列を形態素（意味を持つ最小の単位）に分割して、個々の形態素に対して形態情報（品詞、活用など）を付加していた。

【0003】この形態素を解析する形態素解析システムは、入力の日本語文字列を単語に分解するものである。この処理において、形態素解析システムは、分割される可能性のある単語の全ての候補を単語辞書を利用して解析し、単語間を結んだ単語ネットワークを生成し、そのネットワークから評価関数を利用して最適な経路を選択する。

【0004】

【発明が解決しようとする課題】しかしながら、従来の形態素解析システムは、単語ネットワークを生成する場合、単語辞書のみを利用していたが、この単語辞書のみを利用した解析を行う従来の形態素解析では、単語辞書のすべての単語ネットワークと、分割される可能性のある単語の全てとを比較するため非常に時間がかかり、解

析性能が遅くなるという問題点がある。

【0005】本発明はこのような背景の下になされたもので、単語辞書のすべての単語ネットワークと、分割される可能性のある単語の全てとを比較することなく、形態素解析システム用の候補単語ネットワークを高速に生成する形態素解析システム及びその方法を提供する事にある。

【0006】

【課題を解決するための手段】請求項1記載の発明は、入力された文字列を単語に分割する形態素解析システムにおいて、単語及び品詞情報を記憶している辞書記憶部と、この単語記憶部を検索し、前記文字列を単語に分割し、単語ネットワークを生成する単語ネットワーク生成手段と、この単語ネットワーク生成手段により生成された単語ネットワークから、部分単語ネットワークを検出し、キャッシュ記憶部へ格納するキャッシュ記憶格納手段とを具備することを特徴とする。

【0007】請求項2記載の発明は、請求項1記載の形態素解析システムにおいて、入力される文字列を単語に分解するとき、既に解析された単語ネットワークが前記キャッシュ記憶部にある場合、このキャッシュ記憶部から取得し、解析結果の単語とするキャッシュ記憶部検索手段を具備することを特徴とする。

【0008】請求項3記載の発明は、請求項1または請求項2記載の形態素解析システムにおいて、キャッシュ記憶格納手段が入力された文字列を単語に分解した後、前記単語ネットワークの1単語に収束する部分を検出し、この収束する部分で入力文字列を区切り、部分単語ネットワークを構成し、各部分単語ネットワークを前記キャッシュ記憶部に格納することを特徴とする。

【0009】請求項4記載の発明は、において、入力された文字列を単語に分割する形態素解析方法において、単語及び品詞情報を辞書記憶部へ記憶させる辞書記憶過程と、この単語記憶部を検索し、前記文字列を単語に分割し、単語ネットワークを生成する単語ネットワーク生成過程と、この単語ネットワーク生成過程により生成された単語ネットワークから、部分単語ネットワークを検出し、キャッシュ記憶部へ格納するキャッシュ記憶格納過程とを有することを特徴とする。

【0010】請求項5記載の発明は、請求項4記載の形態素解析方法において、入力される文字列を単語に分解するとき、既に解析された単語ネットワークが前記キャッシュ記憶部にある場合、このキャッシュ記憶部から取得し、解析結果の単語とするキャッシュ記憶部検索過程を有することを特徴とする。

【0011】請求項6記載の発明は、請求項4または請求項5記載の形態素解析システムにおいて、キャッシュ記憶格納過程が入力された文字列を単語に分解した後、前記単語ネットワークの1単語に収束する部分を検出し、この収束する部分で入力文字列を区切り、部分単語

ネットワークを構成し、各部分単語ネットワークを前記キャッシュ記憶部に格納することを特徴とする。

【0012】請求項7記載の発明は、入力された文字列を単語に分割する処理をコンピュータに行わせる形態素解析プログラムを記録した記録媒体において、単語及び品詞情報を辞書記憶部へ記憶させる辞書記憶処理と、この単語記憶部を検索し、前記文字列を単語に分割し、単語ネットワークを生成する単語ネットワーク生成処理と、この単語ネットワーク生成処理により生成された単語ネットワークから、部分単語ネットワークを検出し、キャッシュ記憶部へ格納するキャッシュ記憶格納処理とを有することを特徴とする。

【0013】請求項8記載の発明は、請求項7記載の形態素解析プログラムを記録した記録媒体において、入力される文字列を単語に分解するとき、既に解析された単語ネットワークが前記キャッシュ記憶部にある場合、このキャッシュ記憶部から取得し、解析結果の単語とするキャッシュ記憶部検索処理を有することを特徴とする。

【0014】請求項9記載の発明は、請求項7または請求項8記載の形態素解析プログラムを記録した記録媒体において、キャッシュ記憶格納処理が入力された文字列を単語に分解した後、前記単語ネットワークの1単語に収束する部分を検出し、この収束する部分で入力文字列を区切り、部分単語ネットワークを構成し、各部分単語ネットワークを前記キャッシュ記憶部に格納することを特徴とする。

【0015】本発明は、入力された日本語文字列を単語に分割する形態素解析システムにおいて候補ネットワークを生成する場合に、単語に分割する性能を向上する構成を提供するものである。図1において、キャッシュ記憶部31は、以前に単語ネットワーク生成手段22で処理された単語ネットワークからキャッシュ記憶部格納手段23において検出された部分単語ネットワークを記憶している。辞書記憶部32は日本語の単語辞書であり、日本語の単語と品詞情報を記憶している。

【0016】キャッシュ記憶部検索手段21は、入力装置1から単語ネットワーク生成手段22を通して日本語入力文字列を受け取る。例えば「東京都の人口」という文字列が与えられたとする。キャッシュ記憶部検索手段21は、入力文字列の先頭文字で始まり入力文字列に部分的に一致するキャッシュ記憶部31の中の文字列から最も長い文字列を検索し、見つかった場合はその単語ネットワークを取り出し解析結果とする。

【0017】例えば「東京都」という文字列に対する単語ネットワーク（各単語の接続関係）がキャッシュ記憶部31の中に格納されている場合、単語ネットワーク生成手段22で入力文字列を解析する前に、キャッシュ記憶部検索手段21は、キャッシュ記憶部31から入力文字列に対応する単語ネットワークを取得することができ、「東京都」という文字列に対する解析を行わずに済ま

すことができる。

【0018】一方、キャッシュ記憶部31に一致する文字列が存在しない場合は、単語ネットワーク生成手段22が辞書記憶部32を検索して一致する単語を取り出し解析を行いながら単語ネットワークを生成する。そして、キャッシュ記憶部格納手段23では解析された単語ネットワークにおいて部分単語ネットワークを検出しキャッシュ記憶部31に登録する。

【0019】入力文字列から解析された単語ネットワークは、出力装置4に出力される。このようにして、形態素解析システムの候補ネットワークを構成する処理において、すでに解析された単語ネットワークをキャッシュ記憶部31に格納する。そして、入力文字列に対応する単語ネットワークとして、キャッシュ記憶部31に解析結果が存在する場合は、その解析結果を利用することにより、入力文字列に対する形態素解析の性能向上を可能とする。

【0020】

【発明の実施の形態】以下、図面を参照して本発明の実施形態について説明する。図1は本発明の一実施形態による形態素解析システムの構成を示すブロック図である。この図において、一実施形態による形態素解析システムは、キーボード等の入力装置1と、プログラム制御により動作するデータ処理装置2と、情報を記憶する記憶装置3と、ディスプレイ装置や印刷装置等の出力装置4とを含んで構成されている。

【0021】記憶装置3は、キャッシュ記憶部31と、辞書記憶部32とを備えている。キャッシュ記憶部31は、単語ネットワーク生成手段22で処理された単語ネットワークの中から、キャッシュ記憶部格納手段23により検出された単語ネットワークを逐次格納する。辞書記憶部32は、日本語の単語辞書であり、日本語の単語と品詞情報とをあらかじめ記憶している。

【0022】データ処理装置2は、キャッシュ記憶部検索手段21と、単語ネットワーク生成手段22と、キャッシュ記憶部格納手段23とを備えている。キャッシュ記憶部検索手段21は、単語ネットワーク生成手段22を介して得られた入力文字列と、キャッシュ記憶部31に格納された単語ネットワークとを照合し、照合の結果が一致した場合にキャッシュ記憶部31から単語ネットワークを取り出し単語分割の解析結果とする。

【0023】単語ネットワーク生成手段22は、辞書記憶部32を検索し、入力手段1から入力される入力文字列を単語に分割し、単語ネットワークを生成する。キャッシュ記憶部格納手段23は単語ネットワーク生成手段22により生成された単語ネットワークより部分単語ネットワークを検出し、キャッシュ記憶部31に格納する。ここで、部分単語ネットワークとは、単語ネットワークにおいて、ネットワークの枝が一単語に収束する部分で分割した各単語ネットワークである。

【0024】次に、図1および図2を参照し、一実施形態の動作例を説明する。図2は、一実施形態による形態素解析システムの動作例を示すフローチャートである。

入力装置1は、入力される入力文字列を、単語ネットワーク生成手段22に供給する。このとき、キャッシュ記憶部31には、既に解析された文字列に対する単語ネットワークが格納されているとする。また、辞書記憶部32には、日本語の単語と品詞情報とが格納されている。

【0025】ステップA1において、単語ネットワーク生成手段22には、Node集合とNew_node集合とが保持されている。ここで、Node集合には、初期状態において「head」という「Node」を含んでいる。ここで、「head」は、どのような単語とも接続される。

【0026】次に、ステップA2において、単語ネットワーク生成手段22は、Node集合が空かどうかの判定を行う。そして、単語ネットワーク生成手段22は、Node集合が空でないと判定した場合、処理をステップA3へ進める。

【0027】次に、ステップA3において、単語ネットワーク生成手段22は、Node集合を空でないと判定した場合、New_node集合を空集合に初期化する。そして、単語ネットワーク生成手段22は、処理をステップA4へ進める。

【0028】次に、ステップA4において、キャッシュ記憶部検索手段21は、Node集合の各「Node」について、後続する入力文字列と一致するキャッシュ記憶部31上の単語ネットワークの検索を行う。

【0029】次に、ステップA5において、キャッシュ記憶部検索手段21は、後続する入力文字列と、照合の結果一致する単語ネットワークのデータがキャッシュ記憶部31に存在するか否かの判定を行う。そして、キャッシュ記憶部検索手段21は、照合の結果一致する単語ネットワークのデータがキャッシュ記憶部31に存在すると判定した場合、処理をステップA6へ進める。

【0030】次に、ステップA6において、キャッシュ記憶部検索手段21は、照合の結果一致する最も長い単語ネットワークをキャッシュ記憶部31から取得し、取得した単語ネットワークの最左端「Node」と「head」の「Node」とを接続し、単語ネットワークの最右端「Node」を「New_node集合」に追加する。そして、キャッシュ記憶部検索手段21は、処理をステップA7へ進める。

【0031】次に、ステップA7において、単語ネットワーク生成手段22は、取得した単語ネットワークの文字列と後方一致し、後続する入力文字列に照合の結果一致する単語を辞書記憶部32から検索する。

【0032】次に、ステップA8において、単語ネットワーク生成手段22は、後続する入力文字列に照合の結果一致する単語が辞書記憶部32に存在するか否かの判定を行う。そして、単語ネットワーク生成手段22は、

後続する入力文字列に照合の結果一致する単語が辞書記憶部32に存在すると判定した場合、処理をステップA9へ進める。

【0033】次に、ステップA9において、単語ネットワーク生成手段22は、後続する入力文字列に照合の結果一致する単語が辞書記憶部32に存在すると判定した場合、単語ネットワークの後方一致する部分を除いた部分を「親Node」とし、照合の結果一致した単語を「子Node」として接続し、「子Node」を「New_node」集合に追加する。そして、単語ネットワーク生成手段22は、処理をステップA15へ進める。

【0034】一方、ステップA7において、後続する入力文字列に照合の結果一致する単語が辞書記憶部32に存在しないと判定した場合、処理をステップA15へ進める。

【0035】次に、ステップA15において、単語ネットワーク生成手段22は、New_node集合をnode集合に置換する。

【0036】また、ステップA5において、キャッシュ記憶部検索手段21は、照合するデータが存在しない場合、処理をステップA10へ進める。

【0037】次に、ステップA10において、単語ネットワーク生成手段22は、各Nodeについて後続する入力文字列と一致する単語を辞書記憶部32を検索する。

【0038】次に、ステップA11において、単語ネットワーク生成手段22は、各Nodeについて後続する入力文字列に対して、辞書記憶部32に照合の結果一致する単語が存在するか否かの判定を行う。この判定の結果、単語ネットワーク生成手段22は、各Nodeについて後続する入力文字列に対して、辞書記憶部32に照合の結果一致する単語が存在しない場合、処理をステップA15へ進める。一方、単語ネットワーク生成手段22は、各Nodeについて後続する入力文字列に対して、辞書記憶部32に照合の結果一致する単語が存在する場合、処理をステップA12へ進める。

【0039】次に、ステップA12において、単語ネットワーク生成手段22は、ステップA11において辞書記憶部32に照合の結果一致する単語が存在すると判定された場合、照合した単語を「子Node」として「head」の「node」と接続する。

【0040】次に、ステップA13において、単語ネットワーク生成手段22は、入力文字列に対して「子node」が他に「親Node」を持つか否かの判定を行う。この判定の結果、単語ネットワーク生成手段22は、入力文字列が他に「親Node」を持っていると判定した場合、処理をステップA15へ進める。

【0041】次に、ステップA15において、単語ネットワーク生成手段22は、入力文字列が他に「親Node」を持っていると判定した場合、「New_node集合」を「node集合」へ置換する。

【0042】一方、ステップA13において、単語ネットワーク生成手段22は、入力文字列が他に「親Node」を持っていないと判定した場合、処理をステップA14へ進める。

【0043】次に、ステップA14において、単語ネットワーク生成手段22は、入力文字列が他に「親Node」を持っていないと判定した場合、「子Node」を「New_node集合」に追加した後、処理をステップA15へ進める。

【0044】また、ステップA2において、単語ネットワーク生成手段22は、「Node集合」が空であると判定した場合、処理をステップA16へ進める。

【0045】次に、ステップA16において、単語ネットワーク生成手段22は、解析している入力文字列が文末か否かの判定を行う。そして、単語ネットワーク生成手段22は、解析している入力文字列が文末であると判定した場合、処理をステップA17へ進める。

【0046】次に、ステップA17において、キャッシュ記憶格納手段23は、文末と判定された（一文の解析終了が検出された）場合、入力文字列を解析した単語ネットワークから、nodeの接続が収束する部分を検出し、その単語で単語ネットワークを分割し、分割された結果それぞれを部分単語ネットワークとしてキャッシュ記憶部31に格納する。

【0047】一方、ステップA16において、単語ネットワーク生成手段22は、解析している入力文字列が文末でないと判定した場合、処理をステップA18へ進める。

【0048】次に、ステップA18において、単語ネットワーク生成手段22は、入力文字列の次の文字を未登録単語とし、「New_node集合」に追加し、処理をステップA15へ進める。

【0049】次に、図1、図2及び図3を用いて具体的な応用例を用いて説明する。図3は、入力される“東京都市計画の書籍”という日本語文字列の単語ネットワークを示す図である。キャッシュ記憶部31には、既に解析された“東京都”及び“東京”という文字列に対する単語ネットワークが格納されているとする。また、辞書記憶部32には“東京都”、“東京”、“東”、“京都”、“京”、“都”、“都市”、“市”、“京都市”、“都市計画”、“計画”、“の”、“書籍”、“書”、“籍”が登録されているとする。

【0050】例えば、入力文字列が“東京都市計画の書籍”という日本語文字列の場合、キャッシュ記憶部31上に既に“東京都”、“東京”という2つの文字列に対する単語ネットワークが存在し、“東京都”は“東京都”、“東京-都”、“東-京-都”、“東-京都”という単語の分割で構成され、“東京”は“東京”、“東-京”という単語の分割で構成される場合、ステップA6において、最も長い“東京都”という単語ネットワークが照合される。

【0051】そして、最左端「Node」である“東京都”、

「東京」,「東」が「親Node」である「head」と接続され、最右端「Node」である「東京都」「京都」「都」とが「New_node集合」に追加される。また、単語ネットワーク全体の文字列「東京都」に後方一致し、後続する入力文字列に照合する単語が辞書記憶部32から検索され、「New_node集合」に登録される。この場合「京都市」,「都市」,「都市計画」が登録される。この時点の「New_node集合」は、「東京都」,「京都」,「都」,「京都市」,「都市」,「都市計画」である。

【0052】次に、単語ネットワーク生成手段22は、この「New_node集合」を「node集合」に置換し、「Node集合」が空でないため、「New_node集合」を空にして処理を継続する。この状態では、「Node集合」の各「Node」について後続する入力文字列と一致するキャッシュ記憶部31上の単語ネットワークのデータは存在しないので、単語ネットワーク生成手段22は、辞書記憶部32から単語を検索することになる。

【0053】そして、まず、単語ネットワーク生成手段22は、「東京都」に接続する「市」が「New_node集合」に追加され、「親Node」の「東京都」と接続関係でむすばれる。そして、単語ネットワーク生成手段22により、「京都」,「都」,にも「市」が接続される構成であるが、「市」は既に「親Node」を持っているため、「New_node集合」には追加されない。

【0054】同様に、単語ネットワーク生成手段22は、「計画」,「の」を「New_node集合」に追加する。これと同様の処理を繰り返し、文末まで到達すると図3に示すような「東京都市計画の書籍」に対する「Node」ネットワークが作成される。

【0055】また、「書籍」という単語が辞書記憶部32に登録されていない場合は、単語ネットワーク生成手段22が未登録語として処理し、「書」,「籍」という1文字単位の単語として解析される。そして、文末まで解析が終わると、単語ネットワーク生成手段22は、解析された単語ネットワークに対して単語間の接続が1つの単語に収束する部分を検出し、収束する単語間の部分単語ネットワークをキャッシュ記憶部31に格納する。

【0056】このとき、「東京都市計画の書籍」という文に対しては、単語「の」において接続が収束しており、その前後の「東京都市計画」,「書籍」の単語ネットワークがキャッシュ記憶部31に登録される。

【0057】上述したように、一実施形態の形態素解析システムによれば、既に解析され単語に分解された情報を単語ネットワークとしてキャッシュ記憶部31に格納し、キャッシュ記憶部31に解析結果が存在する場合には、キャッシュ記憶部31上の解析結果を用いることにより、単語分割するときの辞書記憶部32のアクセス回数を削減し解析時間を短縮することができるため、日本語文字列の形態素解析を行うとき、必要な単語の候補ネットワークを生成する場合、解析の性能(速度)が向上

する。

【0058】以上、本発明の一実施形態を図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、本発明の要旨を逸脱しない範囲の設計変更等があっても本発明に含まれる。例えば、図4に示す第2の実施形態による形態素解析システムは、データ処理装置5が、図1で示された一実施形態におけるデータ処理装置2の構成に加え、キャッシュ記憶部制御手段51を有し、記憶装置3Aが二次キャッシュ記憶部33を有する点で異なる。

【0059】キャッシュ記憶部31がメモリで構成されている場合、大容量を確保するのが難しいため、ディスク装置で構成される二次キャッシュ記憶部33を搭載する。また、キャッシュ記憶部制御手段51は、キャッシュ記憶部31上からのデータの取得、格納を制御しており、キャッシュ記憶部31にキャッシュ記憶部格納手段23が単語ネットワークを格納する場合、キャッシュ記憶部31に空き領域があれば、その空き領域に格納し、空き領域がない場合、使用頻度の少ない順に単語ネットワークを二次キャッシュ記憶部33へ移動させ、単語ネットワークを格納する。

【0060】さらに、キャッシュ記憶部検索手段21が単語ネットワークを検索する場合、単語ネットワークがキャッシュ記憶部31に存在する場合はキャッシュ記憶部31から取得し、存在しない場合は二次キャッシュ記憶部33を検索し、二次キャッシュ記憶部33に対応する単語ネットワークが存在すると、この単語ネットワークをキャッシュ記憶部31へ移動させ、必要な単語ネットワークを取得する。このとき、キャッシュ記憶部31に空き領域がない場合は、使用頻度の少ない順に単語ネットワークを二次キャッシュ記憶部33へ移動させる。

【0061】上述したように、第2の実施形態の形態素解析システムによれば、既に解析され単語に分解された情報を単語ネットワークとしてキャッシュ記憶部31または二次キャッシュ記憶部33に格納し、キャッシュ記憶部31または二次キャッシュ記憶部33に解析結果が存在する場合には、キャッシュ記憶部31または二次キャッシュ記憶部33上の解析結果を用いることにより、単語分割するときの辞書記憶部32のアクセス回数を削減し、解析時間を短縮することができるため、日本語文字列の形態素解析を行うとき、必要な単語の候補ネットワークを生成する場合、解析の性能(速度)が向上する。

【0062】また、図2に示す各ステップを実現するためのプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録されたプログラムをコンピュータシステムに読み込ませ、実行することにより図形表示処理を行ってもよい。なお、ここでいう「コンピュータシステム」とは、OSや周辺機器等のハードウェアを含むものとする。また、「コンピュータ読み取り

可能な記録媒体」とは、フロッピーディスク、光磁気ディスク、ROM、CD-ROM等の可搬媒体、コンピュータシステムに内蔵されるハードディスク等の記憶装置のことをいう。

【0063】さらに「コンピュータ読み取り可能な記録媒体」とは、インターネット等のネットワークや電話回線等の通信回線を介してプログラムを送信する場合の通信線のように、短時間の間、動的にプログラムを保持するもの、その場合のサーバやクライアントとなるコンピュータシステム内部の揮発性メモリのように、一定時間プログラムを保持しているものも含むものとする。また上記プログラムは、前述した機能の一部を実現するためのものであっても良く、さらに前述した機能をコンピュータシステムにすでに記録されているプログラムとの組み合わせで実現できるものであっても良い。

【0064】

【発明の効果】本発明によれば、入力された文字列を単語に分割する形態素解析システムにおいて、単語及び品詞情報を記憶している辞書記憶部と、この単語記憶手段を検索し、前記文字列を単語に分割し、単語ネットワークを生成する単語ネットワーク生成手段と、この単語ネットワーク生成手段により生成された単語ネットワークから、部分単語ネットワークを検出し、キャッシュ記憶部へ格納するキャッシュ記憶格納手段とを具備するため、既に解析され単語に分解された情報を単語ネットワークとしてキャッシュ記憶部に格納し、キャッシュ記憶

部に解析結果が存在する場合には、キャッシュ記憶部上の解析結果を用いることにより、単語分割するときの辞書記憶部のアクセス回数を削減し解析時間を短縮することができるため、日本語文字列の形態素解析を行うとき、必要な単語の候補ネットワークを生成する場合、解析の性能（速度）が向上する。

【図面の簡単な説明】

【図1】 本発明の一実施形態による形態素解析システムの構成を示すブロック図である。

【図2】 図1に示す形態素解析システムの動作例を説明するフローチャートである。

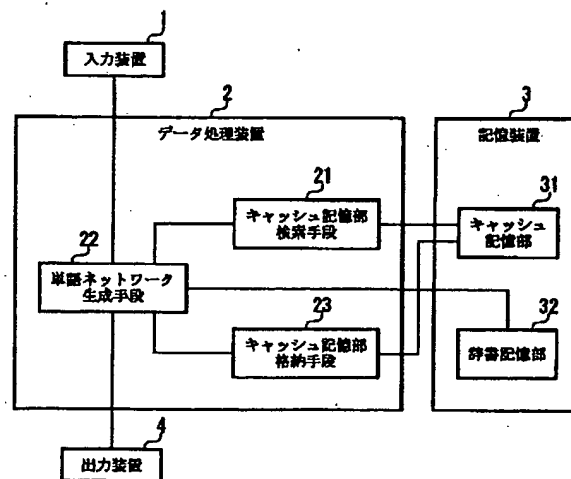
【図3】 入力される「東京都市計画の書籍」という日本語文字列の単語ネットワークを示す図である

【図4】 本発明の第2の実施形態による形態素解析システムの構成を示すブロック図である。

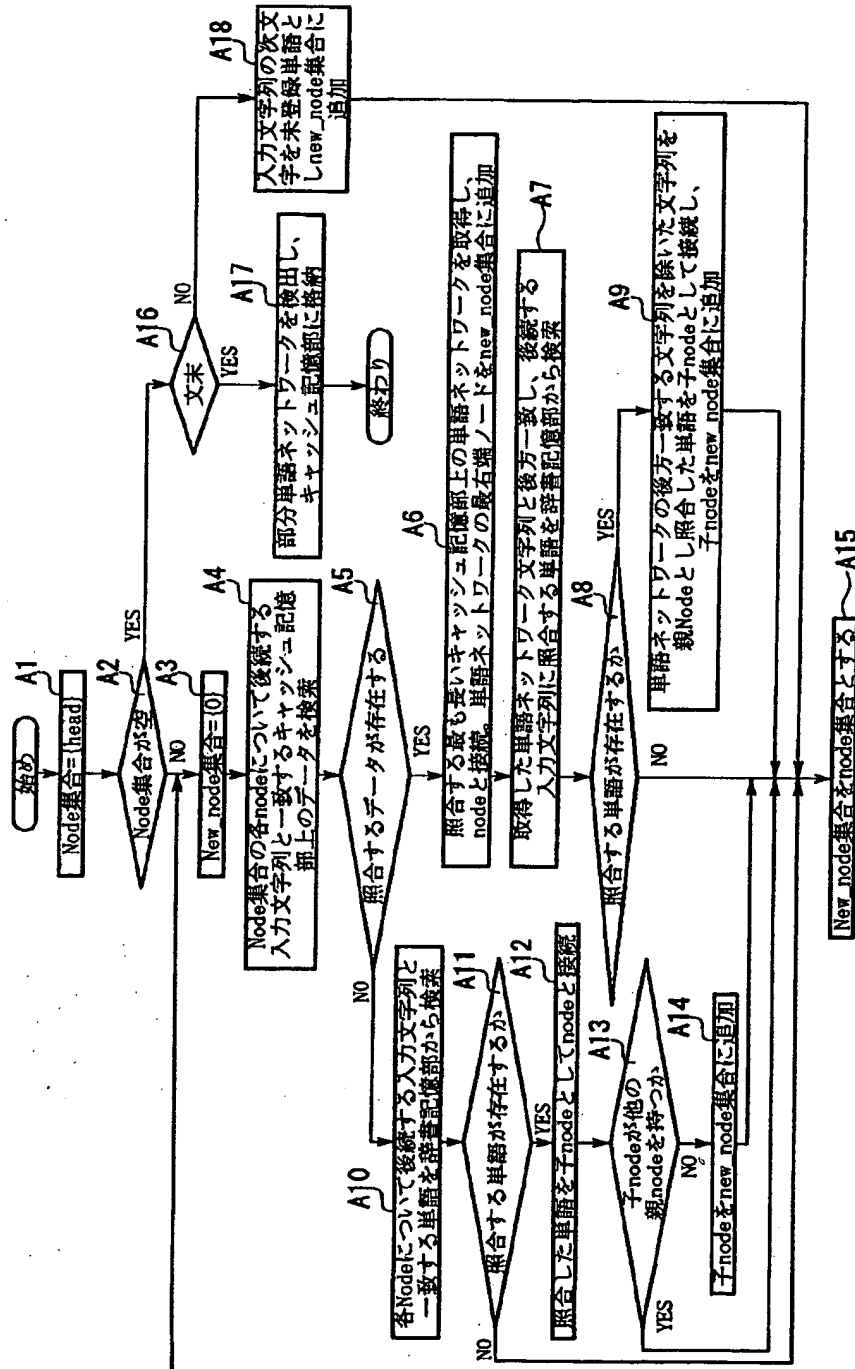
【符号の説明】

- 1 入力装置
- 2、5 データ処理装置
- 3、3A 記憶装置
- 4 出力装置
- 21 キャッシュ記憶部検索手段
- 22 単語ネットワーク生成手段
- 23 キャッシュ記憶部格納手段
- 31 キャッシュ記憶部
- 32 辞書記憶部
- 33 二次キャッシュ記憶部

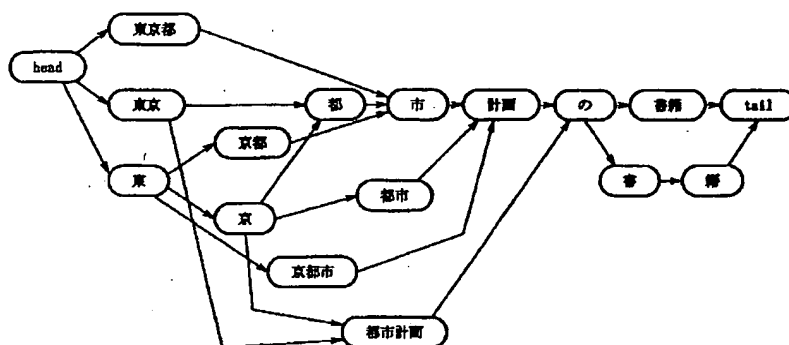
【図1】



【図2】



【図3】



【図4】

